PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C07H 21/04 | A2

(11) International Publication Number: **WO 96/41002**

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/08806

(22) International Filing Date: 6 June 1996 (06.06.96)

(30) Priority Data:

08/487,986	7 June 1995 (07.06.95)	US
08/488,012	7 June 1995 (07.06.95)	US
08/488,013	7 June 1995 (07.06.95)	US
08/487,987	7 June 1995 (07.06.95)	US

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(71) Applicant: GENZYME CORPORATION [–/US]; One Mountain Road, Framingham, MA 01701 (US).

(72) Inventor: SHUBER, Anthony, P.; 11 Grant Street, Milford, MA 01757 (US).

(74) Agent: DUGAN, Deborah, A.; Genzyme Corporation, One Mountain Road, Framingham, MA 01701 (US).

(54) Title: METHODS FOR THE IDENTIFICATION OF GENETIC MODIFICATION OF DNA INVOLVING DNA SEQUENCING AND POSITIONAL CLONING

(57) Abstract

The present invention provides methods for determining the precise location and sequence of genetic alterations and mutations present in a gene of interest. The present invention further provides methods for positional cloning and sequence determination of a gene of interest.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Larvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

METHODS FOR THE IDENTIFICATION OF GENETIC MODIFICATION OF DNA INVOLVING DNA SEQUENCING AND POSITIONAL CLONING

Field of the Invention

This invention pertains to high-throughput methodology that directly identifies previously unidentified sequence alterations in DNA, including specific disease-causing DNA sequences in mammals. The methods of the present invention can be used to identify genetic polymorphisms, to determine the molecular basis for genetic diseases, and to provide carrier and prenatal diagnosis for genetic counseling.

Background of the Invention

The ability to detect alterations in DNA sequences (e.g. mutations and polymorphisms) is central to the diagnosis of genetic diseases and to the identification of clinically significant variants of disease-causing microorganisms. One method for the molecular analysis of genetic variation involves the detection of restriction fragment length polymorphisms (RFLPs) using the Southern blotting technique (Southern, E.M., *J. Mol. Biol.*, 98:503-517, 1975. Since this approach is relatively cumbersome, new methods have been developed, some of which are based on the polymerase chain reaction (PCR). These include: RFLP analysis using PCR (Chehab et al., *Nature*, 329:293-294, 1987; Rommens et al., *Am. J. Hum. Genet.*, 46:395-396, 1990), the creation of artificial RFLPs using primer-specified restriction-site modification (Haliassos et al., *Nuc. Acids Res.*, 17:3606, 1989), allele-specific amplification (ASA) (Newton CR et al., *Nuc. Acids Res.*, 17:2503-2516, 1989), oligonucleotide ligation assay (OLA) (Landergren U et al., *Science* 241:1077-1080, 1988), primer extension (Sokolov BP, *Nuc. Acids Res.*, 18:3671, 1989), artificial introduction of restriction sites (AIRS) (Cohen LB et al., *Nature* 334:119-121, 1988), allele-specific

oligonucleotide hybridization (ASO) (Wallace RB et al., *Nuc. Acids Res.*, 9:879-895, 1981) and their variants. Together with robotics, these techniques for direct mutation and analysis have helped in reducing cost and increasing throughput when only a limited number of mutations need to be analyzed for efficient diagnostic analysis.

These methods are, however, limited in their applicability to complex mutational analysis. For example, in cystic fibrosis, a recessive disorder affecting 1 in 2000-2500 live births in the United States, more than 225 presumed disease-causing mutations have been identified. Furthermore, multiple mutations may be present in a single affected individual, and may be spaced within a few base pairs of each other. These phenomena present unique difficulties in designing clinical screening methods that can accommodate large numbers of sample DNAs.

Shuber et al., *Hum. Mol. Gen.*, 2:153-158, 1993, disclose a method that allows the simultaneous hybridization of multiple oligonucleotide probes to a single target DNA sample. By including in the hybridization reaction an agent that eliminates the disparities in melting temperatures of hybrids formed between synthetic oligonucleotides and target DNA, it is possible in a single test to screen a DNA sample for the presence of different mutations. Typically, more than 100 ASOs can be pooled and hybridized to target DNA; in a second step, ASOs from a pool giving a positive result are individually hybridized to the same DNA. Shuber et al., *Genome Res.* 5:488-93, 1995, disclose a method for multiple allele-specific disease analysis in which multiple ASOs are first hybridized to a target DNA, followed by elution and sequencing of ASOs that hybridize. This method allows the identification of a mutation without the need for many

individual hybridizations involving single ASOs and requires prior knowledge of relevant mutations.

To achieve adequate detection frequencies for rare mutations using the above methods, however, large numbers of mutations must be screened. To identify previously unknown mutations within a gene, other methodologies have been developed, including: single-strand conformational polymorphisms (SSCP) (Orita M et al., *Proc. Natl. Acad. Sci. USA* 86:2766-2770, 1989), denaturing gradient gel electrophoresis (DGGE) (Meyers RM et al., *Nature* 313:495-498, 1985), heteroduplex analysis (HET) (Keen j. et al., *Trends Genet.* 7:5, 1991), chemical cleavage analysis (CCM) (Cotton RGH et al., *Proc. Natl. Acad. Sci. USA* 85:4397-4401, 1988), and complete sequencing of the target sample (Maxam AM et al., *Methods Enzymol.* 65:499-560, 1980, Sanger F. et al., *Proc. Natl. Acad. Sci. USA* 74:5463-5467, 1977). All of these procedures however, with the exception of direct sequencing, are merely screening methodologies. That is, they merely indicate that a mutation exists, but cannot specify the exact sequence and location of the mutation. Therefore, identification of the mutation ultimately requires complete sequencing of the DNA sample. For this reason, these methods are incompatible with high-throughput and low-cost routine diagnostic methods.

Thus, there is a need in the art for a relatively low cost method that allows the efficient analysis of large numbers of DNA samples for the presence of previously unidentified mutations or sequence alterations.

Summary of the Invention

The present invention encompasses high-throughput methods for identifying one or more genetic alterations in

a target sequence present in a first DNA sample. The method is carried out by the steps of:

a) hybridizing the first sample with a second DNA sample not containing genetic alterations to form heteroduplex DNA containing a mismatch region at the site of a genetic alteration(s);

b) cleaving one DNA strand of the heteroduplex in the target sequence to form a single-stranded gap across the site of the alteration;

c) treating the cleaved heteroduplex with a DNA polymerase in the presence of dideoxynucleotides to determine the sequence across the gap; and

d) comparing the nucleotide sequence across the gap with a predetermined cognate wild-type sequence to identify the genetic alteration(s).

In practicing the above-described methods, the first DNA sample containing the target sequence is hybridized under stringent conditions with a second DNA sample not containing the alteration. The hybrids that form contain mismatch regions, which are recognized and endonucleolytically cleaved on one or both sides of the mismatch region by mismatch recognition protein-based systems. When a single endonucleolytic cleavage occurs on only one side of the mismatch region, one or more exonucleases are used to form the single-stranded gap. When endonucleolytic cleavage occurs on both sides of the mismatch region, the single-stranded fragment is released by the action of a helicase to form the single-stranded gap. Determination of the sequence across the gap is achieved in a single step by an enzymatic DNA sequencing reaction using dideoxynucleotides and DNA polymerase I, DNA polymerase III, T4 DNA polymerase, or T7 DNA polymerase.

In an alternate embodiment, the present invention encompasses high-throughput methods for identifying one or more genetic alterations in a target sequence present in a

first DNA sample. This method is carried out by the steps of:

a) hybridizing the first sample with a second DNA sample not containing genetic alterations to form heteroduplex DNA having free ends and containing a mismatch region at the site of a genetic alteration(s);

b) cleaving the DNA at or in the vicinity of the alteration, forming new ends;

c) ligating an oligonucleotide of predetermined sequence to the new ends;

d) determining the nucleotide sequence adjacent to the ligated oligonucleotide; and

e) comparing the nucleotide sequence determined in d) with a predetermined cognate wild-type sequence to identify the genetic alteration(s).

Specific cleavage at or near the alteration is achieved by hybridizing the first DNA sample containing the target sequence with a second DNA sample not containing the alteration, so that heteroduplexes are formed that contain mismatch regions, which can be recognized and cleaved by mismatch recognition systems.

Typically, the first DNA sample comprises genomic DNA from a patient suffering from a genetic disease whose genome does not contain any of the known mutations that cause that disease, and the target sequence comprises a known disease-causing gene. The genetic alterations identified by these methods include additions, deletions, or substitutions of one or more nucleotides.

Mismatch recognition, cleavage, and excision systems useful in practicing the invention include without limitation bacteriophage resolvases, mismatch repair proteins, nucleotide excision repair proteins, chemical modification of mismatched bases followed by excision repair proteins, chemical modification and cleavage, and

combinations thereof, with or without supplementation with exonucleases as required.

The present invention finds application in high-throughput methods for multiplex identification of new mutations or previously unidentified polymorphisms, in which DNAs obtained from a multiplicity of patients are immobilized on a single solid support, followed by one or more of the following steps: hybridization, mismatch recognition, excision, cleavage, ligation, sequencing, and sequence comparison steps as set forth above. Furthermore, multiple specific target sequences can be analyzed simultaneously by amplifying the target sequences prior to immobilization, followed by the steps as set forth above.

In an alternate embodiment, the present invention provides methods for positional cloning of a disease-causing gene. Invention methods are carried out using the following steps:

a) hybridizing a first DNA sample derived from an individual suffering from the disease with a second DNA sample derived from a multiplicity of individuals not suffering from the disease, to form hybrids containing mismatch regions at sites at which the sequence of the first DNA sample diverges from the sequence of the second DNA sample;

b) cleaving one DNA strand in the hybrids to form a single-stranded gap across the site of the alteration;

c) determining the nucleotide sequence across the gap;

d) preparing a synthetic oligonucleotide comprising all or part of the nucleotide sequence determined in c); and

e) identifying a DNA clone derived from a cosmid or a P1 library containing the sequence of the synthetic oligonucleotide prepared in d).

In practicing the present invention, mismatch regions are recognized and endonucleolytically cleaved on one or both sides of the mismatch region by mismatch recognition protein-based systems. When a single endonucleolytic cleavage occurs on only one side of the mismatch region, one or more exonucleases are used to form the single-stranded gap. When endonucleolytic cleavage occurs on both sides of the mismatch region, the single-stranded fragment is released by the action of a helicase to form the single-stranded gap. Determination of the sequence across the gap is achieved in a single step by an enzymatic DNA sequencing reaction using dideoxynucleotides and DNA polymerase I, DNA polymerase III, T4 DNA polymerase, or T7 DNA polymerase.

The present invention further provides alternative methods for positional cloning of a gene of interest. These methods are carried out by:

a) hybridizing a first DNA sample derived from an individual displaying a given phenotype with a second DNA sample derived from one or more individuals not displaying the phenotype, to form heteroduplex DNA having free ends and containing a mismatch region at sites at which the sequence of the first DNA sample diverges from the sequence of the second DNA sample;

b) blocking the free ends on the hybrids formed in a);

c) cleaving one or both DNA strands within or adjacent to the mismatch regions to form new ends;

d) ligating a single-stranded oligonucleotide of predetermined sequence to the new ends formed in c);

e) determining the nucleotide sequence adjacent to the ligatedpredetermined sequence;

f) preparing a synthetic oligonucleotide comprising all or part of the nucleotide sequence determined in e); and

g) identifying a DNA clone derived from a cosmid or a P1 library containing the sequence of the synthetic oligonucleotide prepared in f).

As used herein, positional cloning refers to a process by which a previously unknown disease-causing gene is localized and identified.

The genetic alterations identified by invention methods include additions, deletions, or substitutions of one or more nucleotides. Mismatch recognition, cleavage, and excision systems useful in practicing the invention include without limitation mismatch repair proteins, nucleotide excision repair proteins, bacteriophage resolvases, chemical modification of mismatched bases followed by excision repair proteins, and combinations thereof, with or without supplementation with exonucleases as required.

Detailed Description of the Invention

The present invention encompasses high-throughput methods for identifying specific target sequences in DNA isolated from a patient. As used herein, the term high-throughput refers to a system for rapidly assaying large numbers of DNA samples at the same time. The methods are applicable when one or more genes or genetic loci are targets of interest. The specific sequences typically contain one or more sequence alterations relative to wild-type DNA, including additions, deletions, or substitutions of one or more nucleotides.

In practicing the methods of the present invention, the first DNA sample containing the target sequence is hybridized with a second sample of DNA (or a pool of DNA samples) containing one or more wild-type versions of the targeted gene. The methods of the present

invention take advantage of the physico-chemical properties of DNA hybrids between almost-identical (but not completely identical) DNA strands (i.e., heteroduplexes). When a sequence alteration is present, the heteroduplexes contain a mismatch region that is embedded in an otherwise perfectly matched hybrid. According to the present invention, mismatch regions are formed under controlled conditions and are chemically and/or enzymatically modified; the sequences adjacent to, and including, the mismatch are then determined. Depending upon the mismatch recognition method used, the mismatch region may comprise any number of bases, preferably from 1 to about 1000 bases.

The methods of the invention can be employed to identify specific disease-causing mutations in individual patients (when the gene or genes responsible for the disease are known) or previously unidentified polymorphisms and for positional cloning to identify new genes.

In a preferred embodiment, the specific DNA sequence comprises a portion of a particular gene or genetic locus in the patient's genomic DNA known to be involved in a pathological condition or syndrome. Non-limiting examples of genetic syndromes include cystic fibrosis, sickle-cell anemia, thalassemias, Gaucher's disease, adenosine deaminase deficiency, alpha1-antitrypsin deficiency, Duchenne muscular dystrophy, familial hypercholesterolemia, fragile X syndrome, glucose-6-phosphate dehydrogenase deficiency, hemophilia A, Huntington disease, myotonic dystrophy, neurofibromatosis type 1, osteogenesis imperfecta, phenylketonuria, retinoblastoma, Tay-Sachs disease, and Wilms tumor (Thompson and Thompson, *Genetics in Medicine, 5th Ed.*).

In another embodiment, the specific DNA sequence comprises part of a particular gene or genetic locus that may not be known to be linked to a particular disease, but

in which polymorphism is known or suspected. For example, obesity may be linked with variations in the apolipoprotein B gene, hypertension may be due to genetic variations in sodium or other transport systems, aortic aneurysms may be linked to variations in α-haptoglobin and cholesterol ester transfer protein, and alcoholism may be related to variant forms of alcohol dehydrogenase and mitochondrial aldehyde dehydrogenase. Furthermore, an individual's response to medicaments may be affected by variations in drug modification systems such as cytochrome P450s, and susceptibility to particular infectious diseases may also be influenced by genetic status. Finally, the methods of the present invention can be applied to HLA analysis for identity testing.

In yet another embodiment, the specific DNA sequence comprises part of a foreign genetic sequence e.g. the genome of an invading microorganism. Non-limiting examples include bacteria and their phages, viruses, fungi, protozoa, and the like. The present methods are particularly applicable when it is desired to distinguish between different variants or strains of a microorganism in order to choose appropriate therapeutic interventions.

1. PREPARATION OF HETERODUPLEXES

In accordance with the present invention, the target sequence is contained within a sample of DNA isolated from an animal or human patient. This DNA may be obtained from any cell source or body fluid. Non-limiting examples of cell sources available in clinical practice include blood cells, buccal cells, cervicovaginal cells, epithelial cells from urine, fetal cells, or any cells present in tissue obtained by biopsy. Body fluids include blood, urine, cerebrospinal fluid, and tissue exudates at the site of infection or inflammation. DNA is extracted

from the cell source or body fluid using any of the numerous methods that are standard in the art. It will be understood that the particular method used to extract DNA will depend on the nature of the source. The preferred amount of DNA to be extracted for analysis of human genomic DNA is at least 5 pg (corresponding to about 1 cell equivalent of a genome size of 4 x 10^9 base pairs). In some applications, such as, for example, detection of sequence alterations in the genome of a microorganism, variable amounts of DNA may be extracted.

Once extracted, the sample DNA containing the target sequence may be employed in the present invention without further manipulation. Preferably, one or more specific regions present in the sample DNA may be amplified. In this case, the amplified regions are specified by the choice of particular flanking sequences for use as primers. Amplification at this step provides the advantage of increasing the concentration of specific sequences within the sample DNA population. The length of DNA sequence that can be amplified ranges from 80 bp to up to 30 kbp (Saiki et al., 1988, *Science*, 239:487). Furthermore, the use of amplification primers that are modified by, e.g., biotinylation, allows the selective incorporation of the modification into the amplified DNA.

In one embodiment, the first DNA containing the target sequence, with or without prior amplification of particular sequences, is bound to a solid-phase matrix. This allows the simultaneous processing and screening of a large number of patient or first DNA samples. Non-limiting examples of matrices suitable for use in the present invention include nitrocellulose or nylon filters, glass beads, magnetic beads coated with agents for affinity capture, treated or untreated microtiter plates, and the like. It will be understood by a skilled practitioner that the method by which the DNA is bound to the matrix

will depend on the particular matrix used. For example, binding to nitrocellulose can be achieved by simple adsorption of DNA to the filter, followed by baking the filter at 75-80°C under vacuum for 15 min. to 2h. Alternatively, charged nylon membranes can be used that do not require any further treatment of the bound DNA. Beads and microtiter plates that are coated with avidin can be used to bind DNA that has had biotin attached (via e.g. the use of biotin-conjugated primers). In addition, antibodies can be used to attach DNA to any of the above solid supports by coating the surfaces with the antibodies and incorporating an antibody-specific hapten into the DNA. In a preferred embodiment, DNA that has been amplified using biotinylated primers is bound to streptavidin-coated beads (Dynal, Inc., Milwaukee, WI).

In practicing the present invention, the untreated or amplified first DNA, preferably bound to a solid-phase matrix, is hybridized with a second DNA sample under conditions that favor the formation of mismatch loops. The second DNA sample preferably comprises one or more "wild-type" version(s) of the target sequence. As used herein, a "wild-type" version of a gene is one prevalent in the general population that is not associated with disease (or with any discernable phenotype) and is thus carried by "normal" individuals. In the general population, wild-type genes may include multiple prevalent versions, which contain alterations in sequence relative to each other that cause no discernable pathological effect; these variations are designated "polymorphisms" or "allelic variants". Most preferably, a mixture of DNAs from "normal" individuals is used for the second DNA sample, thus providing a mixture of the most common polymorphisms. This insures that, statistically, hybrids formed between the first and second DNA sample will be perfectly matched except in the region of the mutation, where discrete mismatch regions will form. In some applications, it is

desired to detect polymorphisms; in these cases, appropriate sources for the second DNA sample will be selected accordingly. Depending upon what method is used subsequently to detect mismatches, the wild-type DNA may also be chemically or enzymatically modified, e.g., to remove or add methyl groups.

Hybridization reactions according to the present invention are performed in solutions ranging from about 10 mM NaCl to about 600 mM NaCl, at temperatures ranging from about 37°C to about 65°C. It will be understood that the stringency of a hybridization reaction is determined by both the salt concentration and the temperature; thus, a hybridization performed in 10 mM salt at 37°C may be of similar stringency to one performed in 500 mM salt at 65°C. For the purposes of the present invention, any hybridization conditions may be used that form perfect hybrids between precisely complementary sequences and mismatch loops between non-complementary sequences in the same molecules. Preferably, hybridizations are performed in 600 mM NaCl at 65°C. Following the hybridization step, DNA molecules that have not hybridized to the first DNA sample are removed by washing under stringent conditions, e.g., 0.1X SSC at 65°C.

The hybrids formed by the hybridization reaction may then be treated to block any free ends so that they cannot serve as substrates for further enzymatic modification such as, e.g., by RNA ligase. Suitable blocking methods include without limitation removal of 5' phosphate groups, homopolymeric tailing of 3' ends with dideoxynucleotides, and ligation of modified double-stranded oligonucleotides to the ends of the duplex.

2. MISMATCH RECOGNITION AND CLEAVAGE

In the next step, the hybrids are treated so that one or both DNA strands are cleaved within, or in the vicinity of, the mismatch region. Depending on the method used for mismatch recognition and cleavage (see below), cleavage may occur at some predetermined distance from either boundary of the mismatch region, and may occur on the wild-type or mutant strand. The "vicinity" of the mismatch as used herein thus encompasses from 1 to 2000 bases from the borders of the mismatch. Non-limiting examples of mismatch recognition and cleavage systems suitable for use in the present invention include mismatch repair proteins, nucleotide excision repair proteins, bacteriophage resolvases, chemical modification, and combinations thereof. These embodiments are described below.

In general, the mismatch recognition and/or modification proteins necessary for each embodiment described below are isolated using methods that are well known to those skilled in the art. Preferably, when the sequence of a protein is known, the protein-coding region of the relevant gene is isolated from the source organism by subjecting genomic DNA of the organism to amplification using appropriate primers. The isolated protein-coding DNA sequence is cloned into commercially available expression vectors that, e.g., insert an amino acid "purification tag" at either the amino- or carboxyterminus of the recombinant protein. The recombinant expression vector is then introduced into an appropriate host cell (e.g., *E. coli*), and the protein is recovered from the cell lysate by affinity chromatography that recognizes the "tag". For example, the bacterial expression vector pQiex12 is used to express proteins with a polyhistidine tag, allowing purification of the recombinant product by a single step of chromatography on Ni-Sepharose (QiaGen, Chatsworth, CA).

Other methods involve the expression of recombinant proteins carrying glutathione-S-transferase sequences as tags, allowing purification of the recombinant products on glutathione affinity columns (Pharmacia Biotech, Uppsala, Sweden). If necessary, proteins containing purification tags are then treated so as to remove the tag sequences. Alternatively, the protein may be isolated from its cell of origin using standard protein purification techniques well-known in the art, including, e.g., molecular sieve, ion-exchange, and hydrophobic chromatography; and isoelectric focusing. "Isolation" as used herein denotes purification of the protein to the extent that it can carry out its function in the context of the present invention without interference from extraneous proteins or other contaminants derived from the source cells.

The mismatch recognition and modification proteins used in practicing the present invention may be derived from any species, from *E. coli* to humans, or mixtures thereof. Typically, functional homologues for a given protein exist across phylogeny. A "functional homologue" of a given protein as used herein is another protein that can functionally substitute for the first protein, either *in vivo* or in a cell-free reaction.

Mismatch repair proteins:

A number of different enzyme systems exist across phylogeny to repair mismatches that form during DNA replication. In *E. coli*, one system involves the MutY gene product, which recognizes A/G mismatches and cleaves the A-containing strand (Tsai-Wu et al., *J. Bacteriol.* 178:1902,1991). Another system in *E. coli* utilizes the coordinated action of the MutS, MutL, and MutH proteins to recognize errors in newly-synthesized DNA strands specifically by virtue of their transient state of undermethylation (prior to their being acted upon by dam

methylase in the normal course of replication). Cleavage typically occurs at a hemimethylated GATC site within 1-2 kb of the mismatch, followed by exonucleolytic cleavage of the strand in either a 3'-5' or 5'-3' direction from the nick to the mismatch. *In vivo*, this is followed by re-synthesis involving DNA polymerase III holoenzyme and other factors (Cleaver, *Cell*, 76:1-4, 1994).

Mismatch repair proteins for use in the present invention may be derived from *E. coli* (as described above) or from any organism containing mismatch repair proteins with appropriate functional properties. Non-limiting examples of useful proteins include those derived from *Salmonella typhimurium* (MutS, MutL); *Streptococcus pneumoniae* (HexA, HexB); *Saccharomyces cerevisiae* ("all-type", MSH2, MLH1, MSH3); Schizosaccharomyces pombe (SWI4); mouse (rep1, rep3); and human ("all-type", hMSH2, hMLH1, hPMS1, hPMS2, duc1). Preferably, the "all-type" mismatch repair system from human or yeast cells is used (Chang et al., *Nuc. Acids Res.* 19:4761, 1991; Yang et al., *J. Biol. Chem.* 266:6480,1991). In a preferred embodiment, heteroduplexes formed between patients' DNA and wild-type DNA as described above are incubated with human "all-type" mismatch repair activity that is purified essentially as described in International Patent Application WO/93/20233.

Incubations are performed in, e.g., 10mM Tris-HCl pH 7.6, 10mM $ZnCl_2$, 1mM dithiothreitol, 1mM EDTA and 2.9% glycerol at 37°C for 1-3 hours. In another embodiment, purified MutS, MutL, and MutH are used to cleave mismatch regions (Su et al., *Proc. Natl. Acad. Sci.USA* 83:5057,1986; Grulley et al., *J. Biol. Chem.* 264:1000,1989).

Nucleotide excision repair proteins:

In *E. coli*, four proteins, designated UvrA, UvrB, UvrC, and UvrD, interact to repair nucleotides that are

damaged by UV light or otherwise chemically modified (Sancar, *Science* 266:1954, 1994), and also to repair mismatches (Huang et al., *Proc. Natl. Acad.Sci.USA* 91:12213, 1994). UvrA, an ATPase, makes an A_2B_1 complex with UvrB, binds to the site of the lesion, unwinds and kinks the DNA, and causes a conformational change in UvrB that allows it to bind tightly to the lesion site. UvrA then dissociates from the complex, allowing UvrC to bind. UvrB catalyzes an endonucleolytic cleavage at the fifth phosphodiester bond 3' from the lesion; UvrC then catalyzes a similar cleavage at the eighth phosphodiester bond 5' from the lesion. Finally, UvrD (helicase II) releases the excised oligomer. *In vivo*, DNA polymerase I displaces UvrB and fills in the excision gap, and the patch is ligated.

In one embodiment of the present invention, heteroduplexes formed between patients' DNA and wild-type DNA are treated with a mixture of UvrA, UvrB, UvrC, with or without UvrD. As described above, the proteins may be purified from wild-type *E. coli*, or from *E. coli* or other appropriate host cells containing recombinant genes encoding the proteins, and are formulated in compatible buffers and concentrations. The final product is a heteroduplex containing a single-stranded gap covering the site of the mismatch.

Excision repair proteins for use in the present invention may be derived from *E. coli* (as described above) or from any organism containing appropriate functional homologues. Non-limiting examples of useful homologues include those derived from *S. cerevisiae* (RAD1, 2, 3, 4, 10, 14, and 25) and humans (XPF, XPG, XPD, XPC, XPA, ERCC1, and XPB) (Sancar, *Science* 266:1954, 1994). When the human homologues are used, the excised patch comprises an oligonucleotide extending 5 nucleotides from the 3' end of the lesion and 24 nucleotides from the 5' end of the

lesion. Aboussekhra et al. (*Cell* 80:859, 1995) disclose a reconstituted *in vitro* system for nucleotide excision repair using purified components derived from human cells.

Chemical Mismatch Recognition:

Heteroduplexes formed between patients' DNA and wild-type DNA according to the present invention may be chemically modified by treatment with osmium tetroxide (for mispaired thymidines) and hydroxylamine (for mispaired cytosines), using procedures that are well known in the art (see, e.g., Grompe, *Nature Genetics* 5:111, 1993; and Saleeba et al., *Meth. Enzymol.* 217:288, 1993). In one embodiment, the chemically modified DNA is contacted with excision repair proteins (as described above). The hydroxylamine- or osmium-modified bases are recognized as damaged bases in need of repair, one of the DNA strands is selectively cleaved, and the product is a gapped heteroduplex as above.

Resolvases:

Resolvases are enzymes that catalyze the resolution of branched DNA intermediates that form during recombination events (including Holliday structures, cruciforms, and loops) via recognition of bends, kinks, or DNA deviations (Youil et al., *Proc.Natl.Acad.Sci.USA* 92:87, 1995). For example, Endonuclease VII derived from bacteriophage T4 (T4E7) recognizes mismatch regions of from one to about 50 bases and produces double-stranded breaks within six nucleotides from the 3' border of the mismatch region. T4E7 may be isolated from, e.g., a recombinant *E. coli* that overexpresses gene 49 of T4 phage (Kosak et al.,*Eur. J. Biochem.* 194:779, 1990). Another suitable resolvase for use in the present invention is Endonuclease I of bacteriophage T7 (T7E1), which can be isolated using a

polyhistidine purification tag sequence (Masnal et al., *Nature Genetics* 9:177, 1995).

In a preferred embodiment, heteroduplexes formed between patients' DNA and wild-type DNA as described above are incubated in a 50 µl reaction with 100-3000 units of T4E7 for 1 hour at 37°C.

3. SEQUENCE DETERMINATION

In practicing the present invention, immobilized DNA from a patient is hybridized to wild-type DNA to form mismatch regions and then treated with mismatch repair proteins, excision repair proteins, resolvases, chemical modification and cleavage reagents, or combinations of such agents, to introduce single- or double-stranded breaks at some predetermined location relative to the site of the mismatch regions.

In one embodiment, the introduction of single-stranded breaks at predetermined locations on one or both sides of a mismatch region causes the selective excision of a single-stranded fragment covering the mismatch region. The resulting structure is a gapped heteroduplex in which the gap may be from about 5 to about 2000 bases in length, depending on the mismatch recognition system used.

To determine the nucleotide sequence of the excised region (including the mismatch), the heteroduplexes are incubated with an appropriate DNA polymerase enzyme in the presence of dideoxynucleotides. Suitable enzymes for use in this step include without limitation DNA polymerase I, DNA polymerase III holoenzyme, T4 DNA polymerase, and T7 DNA polymerase. The only requirement is that the enzyme be capable of accurate DNA synthesis using the gapped heteroduplex as a substrate. The presence of dideoxynucleotides, as in a Sanger sequencing reaction,

insures that a nested set of premature termination products will be produced, and that resolution of these products by, e.g., gel electrophoresis will display the DNA sequence across the gap.

In some circumstances, the sequence obtained using this method will correspond to the wild-type strand and not to the patient's DNA in which the mutation is sought. This result is easily accomodated by a second round of sequencing, with or without prior amplification of the relevant DNA region. In this case, the sequence of the mutation is determined using as a template the patient's unmodified DNA in conjunction with sequencing primers derived from the sequence determined in the first round.

In an alternative embodiment of sequence determination, the hybrids formed between the wild-type DNA and the patient's DNA are then dissociated by denaturation, and the wild-type DNA and any cleavage products of the target DNA are removed by washing. The immobilized remaining target DNA is then ligated to a synthetic single-stranded oligonucleotide of predetermined sequence, designated a "ligation oligonucleotide", that serves as a primer for enzymatic DNA sequencing. The oligonucleotide may be from about 15 to about 25 nucleotides in length. A preferred ligation oligonucleotide has the sequence 5'-CAGTAGTACAACTGACCCTTTTGGGACCGC-3'. Ligation is achieved using, e.g., RNA ligase (Pharmacia Biotech, Uppsala, Sweden).

A typical ligation reaction is performed at 37°C for 15 min in a 20 µl reaction containing 50mM Tris-HCl, pH 7.5, 10mM $MgCl_2$, 20mM dithiothreitol, 1mM ATP, 100 µg/ml bovine serum albumin, at least 1 µg immobilized target DNA, a 10-fold molar excess of the ligation oligonucleotide, and 0.1-5.0 units/ml T4 RNA ligase. Following the ligation, unligated oligonucleotides are removed by washing.

The sequence of DNA immediately adjacent to the ligated oligonucleotide is then determined by any method known in the art. In one embodiment, enzymatic sequencing is performed according to the dideoxy Sanger technique, using as a sequencing primer a second oligonucleotide of predetermined sequence that is complementary to the ligation oligonucleotide (Sanger et al., *Proc. Natl. Acad. Sci.USA* 74:5463, 1977). Each microsequencing reaction is then resolved by techniques well-known in the art, including without limitation gel electrophoresis, and the sequence is determined.

In another embodiment, an oligonucleotide complementary to the ligated oligonucleotide is used to prime DNA synthesis using DNA polymerase I in the presence of all four nucleoside triphosphates. The newly synthesized strand is then analyzed using hybridization to oligonucleotide arrays as described in Pease et al., *Proc. Natl. Acad. Sci.USA* 91:5022, 1994.

Identification of a sequence alteration according to the present invention is preferably achieved in a single round of mismatch recognition and cleavage, oligonucleotide ligation, and DNA sequencing. This occurs when the ligated oligonucleotide becomes covalently attached to a) the immobilized truncated target DNA that contains the alteration b) within 10-500 bp of either boundary of the mismatch region. If either of these conditions is not fulfilled, further rounds of sequencing may be required to localize and identify the sequence alteration. It will be understood by those of ordinary skill in the art that sequencing primers for one or more further rounds of sequencing will be dictated by the sequence obtained in the first round (either the same or complementary strands). Without wishing to be bound by theory, it is contemplated that one or two sequencing rounds will reveal the

divergence between a known wild-type sequence and that contained within the DNA of a particular patient (see below).

High-Throughput Applications

The methods of the present invention are particularly suitable for high-throughput analysis of DNA, i.e., the rapid and simultaneous processing of DNA samples derived from a large number of patients. Furthermore, in contrast to other methods for *de novo* mutation detection, the methods of the present invention are suitable for the simultaneous analysis of a large number of genetic loci in a single reaction; this is designated "multiplex" analysis. Therefore, for any one sample or for a multiplicity of samples, the present invention allows the analysis of both intragenic loci (several regions within a single gene) and intergenic loci (several regions within different genes) in a single reaction mixture. The manipulations involved in practicing the methods of the present invention lend themselves to automation, e.g., using multiwell microtiter dishes as a solid support or as a receptacle for, e.g., beads; robotics to perform sequential incubations and washes; and, finally, automated sequencing using commercially available automated DNA sequencers. It is contemplated that, in a clinical context, 500 patient DNA samples can be analyzed within 1-2 days in a cost-effective manner.

Positional Cloning

The methods of the present invention are also suitable for positional cloning of unknown genes that cause pathological conditions or other detectable phenotypes in any organism. "Positional cloning" as used herein denotes a process by which a previously unknown disease-causing gene is localized and identified. For example,

identification of multiplex families in which several members exhibit signs of a genetically-based syndrome often occurs even when the particular gene responsible for the syndrome has not been identified. Typically, the search for the unknown gene involves one or more of the following time- and labor-intensive steps: 1) cytogenetic localization of the gene to a relatively large segment of a particular chromosome; 2) assembly of overlapping cosmid or P1 clones that collectively cover several hundred thousand nucleotides corresponding to the identified chromosomal region; 3) sequencing the clones; and 4) transcript mapping to identify expressed protein-encoding regions of the gene.

The present invention offers an alternative, cost-effective method for localizing a disease-causing gene. Briefly, DNA from affected individuals is hybridized with normal DNA as described above to form mismatch regions at the site of the mutation. Preferably, large regions of DNA corresponding to the chromosomal location are amplified from the patient's genomic DNA prior to inclusion in the hybridization reaction. The hybrids are then treated by any of the methods described above so that mismatch regions are recognized and cleaved, forming gapped heteroduplexes across the mismatch region. Finally, the sequence in the vicinity of the mismatch region is determined.

In this embodiment, determination of even a short sequence in the vicinity of the mismatch facilitates definitive identification of the disease-causing gene. The short sequence that is determined in the first round of sequencing can be used to design oligonucleotide probes for use in screening genomic or cDNA libraries.

Other methods in which the primary sequence information can be used, either alone or in conjunction with library screening, include identification of tissue specific expression, reverse transcription-amplification of

mRNA, and screening of an affected population for genotype/phenotype association. Thus, without wishing to be bound by theory, it is contemplated that a previously unknown gene that causes a disease or other phenotype can be quickly and efficiently identified by these methods.

The following examples are intended to illustrate the present invention without limitation.

Example 1: Preparation of Target DNA

A) Preparation of Sample DNA from Blood

Whole blood samples collected in high glucose ACD Vacutainers™ (yellow top) were centrifuged and the buffy coat collected. The white cells were lysed with two washes of a 10:1 (v/v) mixture of 14mM NH_4Cl and 1mM $NaHCO_3$, their nuclei were resuspended in nuclei-lysis buffer (10mM Tris, pH 8.0, 0.4M NaCl, 2mM EDTA, 0.5% SDS, 500 µg/ml proteinase K) and incubated overnight at 37°C. Samples were then extracted with a one-fourth volume of saturated NaCl and the DNA was precipitated in ethanol. The DNA was then washed with 70% ethanol, dried, and dissolved in TE buffer (10mM Tris-HCl, pH 7.5, 1mM EDTA).

B) Preparation of Sample DNA from Buccal Cells

Buccal cells were collected on a sterile cytology brush (Scientific Products) or female dacron swab (Medical Packaging Corp.) by twirling the brush or swab on the inner cheek for 30 seconds. DNA was prepared as follows, immediately or after storage at room temperature or at 4°C. The brush or swab was immersed in 600 µl of 50mM NaOH contained in a polypropylene microcentrifuge tube and vortexed. The tube, still containing the brush or swab, was heated at 95°C for 5 min, after which the brush or swab was carefully removed. The solution containing DNA was

then neutralized with 60 µl of 1M Tris, pH 8.0, and vortexed again (Mayall et al., *J.Med.Genet.* 27:658, 1990). The DNA was stored at 4°C.

C) Amplification of Target DNA Prior to Hybridization

DNA from patients with CF was amplified by PCR in a Perkin-Elmer Cetus 9600 Thermocycler. Five primer sets were used to simultaneously amplify relevant regions of exons 4, 10, 20, and 21 of the cystic fibrosis transmembrane conductance regulator (CFTR) gene (Richards et al., *Hum. Mol.Gen.* 2:159, 1993). The 50 µl PCR reaction mix contained the following components: 0.2-1 µg CF patient DNA, 10mM Tris pH 8.3, 50mM KCl, 1.5mM $MgCl_2$, 0.01% (w/v) gelatin, 200µM of each deoxynucleotide triphosphate, 0.4µM of each amplification primer, and 2.5 units of Taq polymerase. An initial denaturation was performed by incubation at 94°C for 20 seconds, followed by 28 cycles of amplification, each consisting of 10 seconds at 94°C, 10 seconds at 55°C, 10 seconds at 74°C, and a final soak at 74°C for 5 min. Following amplification, 8 µl of the PCR products were electrophoresed in a 2% agarose gel to verify the presence of all five products.

D) Binding of DNA to a Solid Matrix:

For binding of amplified DNA to a solid support, the amplification reactions described above are performed in the present of biotinylated primers. The biotinylated products are then incubated with Dynabeads®M-280 Streptavidin (Dynal) in a solution containing 10 mM Tris HCl, pH 7.5, 1 mM EDTA, 2M NaCl, and 0.1% Tween-20 for 15-30 minutes at 48°C.

Example 2: Hybridization of target DNA and wild-type DNA

A) Preparation of wild-type DNA:

DNA is prepared from blood or buccal cells of healthy individuals as described in Example 1. A representative "wild-type" DNA sample is prepared by combining and thoroughly mixing DNA samples derived from 10-200 individuals.

B) Hybridization Reaction:

Hybridizations are carried out in microtiter dishes containing bead-immobilized DNA prepared as in Example 1D above. The hybridization solution contains approximately 500 µg/ml wild-type DNA (prepared as in Example 2A above) and approximately 50 µg/ml amplified immobilized target DNA (prepared as in Example 1) in 10mM Tris HCl pH 7.5 - 650mM NaCl. The reaction mixtures are heated at 90°C for 3 minutes, after which hybridizations are allowed to proceed for 1 hour at 65°C. The hybridization solution is then removed and the beads are washed three times in 0.1x SSC at 65°C.

C) Blocking of free ends:

The beads containing DNA:DNA hybrids prepared as described above are treated so that free ends become blocked and no longer accessible to modification by, e.g., RNA ligase. The wells are incubated in 100 µl of a solution containing 0.4M potassium cacodylate, 50 mM Tris HCl, pH 6.9, 4 mM dithiothreitol, 1 mM $CoCl_2$, 2mM ddGTP, 500 µg/ml bovine serum albumin, and 2 units of terminal transferase for 15 minutes at 37°C.

Example 3: Mismatch recognition, cleavage, and sequencing

A) In one embodiment of the present invention, four identical reactions mixtures, each containing 50 µl beads to which DNA hybrids prepared as described in Example 2 are bound, are incubated with 2 µl of a 10X T4 Polymerase buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 1mM dithiothreitol, and 1 mg/ml bovine serum albumin); 16 µl water; 1 µl T4 endonuclease 7 (250-3000 units, obtained as described in Kosak et al., *Eur.J.Biochem.* 194:779, 1990); and 1 µl T7 DNA polymerase (3 units). The reaction is allowed to proceed for 1-10 minutes at 37°C.

9 µl of a "termination mix" is then added to each reaction. "Termination mix" contains 8 µM of a single ddNTP (i.e., ddGTP, ddATP, ddTTP, or ddCTP) and 80 µM of all four dNTPs, one of which is labelled with a radioactive or fluorescent label. In addition, 1 µl of 10X T4 polymerase buffer is added, and the reaction is allowed to proceed for 5 minutes at 37°C.

The reaction mix is removed and the beads are washed three times with 100 µl TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA). Finally, the beads are resuspended in 6 µl gel loading buffer (95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% Xylene Cyanol FF). The buffer is removed from the beads and loaded on a 6% denaturing polyacrylamide DNA sequencing gel.

B) Alternatively, 50 µl beads containing DNA hybrids prepared as described in Example 2 are incubated with 500 units of T4 endonuclease 7 in a solution containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, and 1 mM dithiothreitol for 30 minutes at 37°C. T4 endonuclease 7 is obtained as described in Kosak et al., *Eur. J. Biochem.* 194:779,1990.

After the incubation, the beads are heated to 90°C for three minutes, after which the solution is quickly removed and replaced with prewarmed TE, and the beads are washed three times with TE at room temperature. This procedure effectively denatures DNA:DNA hybrids and removes wild-type DNA strands.

Example 4: Mismatch recognition and cleavage using chemical mismatch cleavage

In one embodiment of the present invention, microtiter wells prepared as described in Examples 1 and 2 above are treated sequentially with hydroxylamine and osmium tetroxide.

A) Hyroxylamine treatment:

Hydroxylamine (obtained from Aldrich, Milwaukee, WI) is dissolved in distilled water, and the pH is adjusted to 6.0 with diethylamine (Aldrich) so that the final concentration is about 2.5 M. 200 µl of the solution are incubated within the wells at 37°C for 2 hours. The reaction is stopped by replacing the hydroxylamine solution with an ice-cold solution containing 0.3 M sodium acetate, 0.1mM EDTA, pH5.2, and 25 µg/ml yeast tRNA (Sigma, St. Louis, MO). The wells are then washed in an ice-cold solution of 10mM Tris-HCl, pH 7.7, 1mM EDTA prior to osmium tetroxide treatment.

B) Osmium tetroxide treatment:

Osmium tetroxide (Aldrich) is dissolved in 10mM Tris-HCl, pH 7.7, 1mM EDTA, and 1.5% (v/v) pyridine to a concentration of 4% (w/v). The wells are incubated with this solution for 2 hours at 37°C. The reaction is stopped by replacing the osmium tetroxide solution with an

ice-cold solution containing 0.3 M sodium acetate, 0.1mM EDTA, pH5.2, and 25 µg/ml yeast tRNA.

C) Piperidine cleavage:

Chemical cleavage of the C and T bases that react with hydroxylamine or osmium tetroxide is achieved by incubating the dishes with 1M piperidine at 90°C for 30 min. The wells are then washed extensively with distilled water.

Example 5: Sequencing of mismatch regions

Immobilized DNAs prepared as described in Examples 1 and 2 above and subjected to mismatch recognition and cleavage (as described in Examples 3B or 4 above or by other methods) are incubated with a single-stranded oligonucleotide having the sequence 5'-CAGTAGTACAACTGACCCTTTTGGGACCGC-3' under conditions in which efficient ligation of the oligonucleotide to free 5' ends is achieved. The oligonucleotide and immobilized DNA are combined in a solution containing 50 mM Tris HCl, pH 7.5, 10 mM $MgCl_2$, 20 mM dithiothreitol, 1 mM ATP, and 100 µg/ml bovine serum albumin, after which RNA ligase (Pharmacia Biotech, Uppsala, Sweden) is added to the solution to achieve a final enzyme concentration of 0.1-5.0 U/ml. The reaction is allowed to proceed at 37°C for 15 min. Following the ligation reaction, the solution is removed, and the wells are washed with distilled water.

DNA sequencing is then performed using the Sanger method (Sanger et al., *Proc.Natl.Acad.Sci.USA* 74:5463, 1977).

Example 6: Positional cloning of a disease-causing gene

The experiments described below are performed to rapidly localize and sequence a genomic region corresponding to a disease-causing gene.

A multiplex family in which a genetic disease is expressed is identified using standard clinical indicators. DNA samples are obtained from affected and unaffected individuals as described in Example 1 above; if by patterns of transmission the disease appears to be an autosomal recessive syndrome, DNA samples are obtained from those individuals presumptively heterozygous for the disease gene.

In one embodiment, all DNA samples are subjected to mismatch analysis by hybridization to wild-type DNA as described in Example 2 above. The hybrids are then treated with mismatch repair proteins to form a gapped heteroduplex, and the sequence across the gap is determined as described in Example 3A above.

In an alternative embodiment, all DNA samples are subjected to mismatch analysis by hybridization to wild-type DNA as described in Example 2 above. The hybrids are then treated with T4 endonuclease 7 as described in Example 3B above. Finally, an oligonucleotide having the sequence 5'-CAGTAGTACAACTGACCCTTTTGGGACCGC-3' is ligated to the cleaved hybrids using RNA ligase, and the products are subjected to enzymatic DNA sequencing as described in Example 5 above.

The sequences obtained from unaffected, affected, and presumptively heterozygous family members are compared with each other and with available sequence databases, using, for example, Sequencher (Gene Codes, Ann Arbor, MI) and Assembly Lign (Kodak, New Haven, CT) The sequences are

also serve as the basis for design of oligonucleotide probes, which are chemically synthesized and used to probe human genomic DNA libraries.

What is claimed is:

1. A method for identifying one or more genetic alterations in a target sequence present in a first genomic DNA sample, which comprises:

a) hybridizing said DNA sample with a second DNA sample, wherein said second sample does not contain the alteration(s), to form heteroduplex DNA containing a mismatch region at the site of an alteration(s);

b) cleaving one strand of said heteroduplex in the target sequence to form a single-stranded gap across the site of said alteration(s);

c) treating said cleaved heteroduplex with a DNA polymerase in the presence of dideoxynucleotides to determine the nucleotide sequence across said gap; and

d) comparing said nucleotide sequence with a predetermined cognate wild-type sequence to identify said genetic alteration(s).

2. The method of claim 1, 16, 24 or 32, wherein the alterations are selected from the group consisting of additions, deletions, and substitutions of one or more nucleotides and combinations thereof.

3. The method of claim 1, 16, 24 or 32, wherein said target sequence is amplified prior to the hybridizing step.

4. The method of claim 1, 16 or 24, wherein the first DNA sample is immobilized on a solid support prior to the hybridizing step.

5. The method of claim 4, wherein the solid support is selected from the group consisting of nitrocellulose filter, nylon filter, glass beads, and plastic.

6. The method of claim 1, 21, 24 or 35, wherein said cleaving step comprises exposing said heteroduplex DNA to one or more resolvase proteins under conditions appropriate for mismatch recognition and cleavage.

7. The method of claim 6, wherein the resolvases are selected from the group consisting of T4 endonuclease 7 and T7 endonuclease 1.

8. The method of claim 1 or 21, wherein said DNA polymerase is selected from the group consisting of DNA polymerase I, DNA polymerase III, T7 DNA polymerase, and T4 DNA polymerase.

9. The method of claim 1, 21, 24 or 35, wherein said cleaving step comprises exposing said heteroduplex DNA to one or more mismatch repair proteins under conditions appropriate for mismatch recognition, cleavage, and excision.

10. The method of claim 9, wherein the one or more mismatch repair proteins comprise *Escherichia coli* proteins MutS, MutL, MutH, and MutU, or functional homologues thereof.

11. The method of claim 10, wherein the functional homologues are derived from species selected from the group consisting of *Salmonella typhimurium, Streptococcus pneumoniae, Saccharomyces cerevisiae, Schizosaccharomyces pombe*, mouse and human.

12. The method of claim 1, 21, 24 or 35, wherein said cleaving step comprises exposing said heteroduplex DNA to a mixture of nucleotide excision repair

proteins under conditions appropriate for mismatch recognition, cleavage, and excision.

13. The method of claim 12, wherein the mixture comprises *E. coli* proteins UvrA, UvrB, UvrC, and UvrD, or functional homologues thereof.

14. The method of claim 13, wherein the functional homologues are derived from species selected from the group consisting of *Saccharomyces cerevisiae* and human.

15. The method of claim 1 or 24, further comprising determining the complement of said nucleotide sequence using said first DNA as a template.

16. A method for identifying one or more genetic alterations in a target sequence present in a first genomic DNA sample, which comprises:

a) hybridizing the first DNA sample with a second DNA sample, wherein said second sample does not contain the alteration(s), to form heteroduplex DNA containing a mismatch region at the site of an alteration(s);

b) treating said heteroduplex DNA with a mixture of T4 endonuclease 7 and DNA polymerase I in the presence of dideoxynucleotides to form premature termination products;

c) resolving said termination products to determine the nucleotide sequence in the vicinity of the mismatch region; and

d) comparing said nucleotide sequence with a predetermined cognate wild-type sequence to identify said alteration(s).

17. A method for multiplex identification of one or more mutation(s) in a DNA, the method comprising:

a) immobilizing one or more first DNA samples on a solid support;

b) hybridizing said immobilized sample(s) with a second DNA sample, wherein said second sample does not contain the mutation(s), to form heteroduplex DNA containing a mismatch region at the site of a mutation(s);

c) cleaving one or both strands of said heteroduplex adjacent to said mismatch region to form a gap at the site of said mutation(s);

d) treating said cleaved heteroduplex with a DNA polymerase in the presence of dideoxynucleotides to determine the nucleotide sequence across said gap using enzymatic DNA sequencing; and

e) comparing said nucleotide sequence(s) with one or more predetermined cognate wild-type sequences to identify said mutation(s).

18. The method of claim 1, 16, 17, 24, 33 or 34, wherein the DNA samples are denatured prior to hybridization.

19. The method of claim 17, 33 or 34, wherein the first DNA sample is amplified prior to immobilization.

20. A method for identifying one or more genetic alterations in a target sequence present in a genomic DNA sample, which comprises:

a) denaturing said DNA;

b) reannealing said DNA to form heteroduplex DNA containing a mismatch region at the site of an alteration(s);

c) cleaving one strand of said heteroduplex in said target sequence to form a single-stranded gap across the site of said alteration(s);

d) treating said cleaved heteroduplex with a DNA polymerase in the presence of dideoxynucleotides to determine the nucleotide sequence across said gap; and

e) comparing said nucleotide sequence with a predetermined cognate wild-type sequence to identify said alteration(s).

21. A method for positional cloning of a gene of interest, the method comprising:

a) hybridizing a first DNA sample derived from an individual displaying a given phenotype with a second DNA sample, wherein said second DNA sample is derived from one or more individual(s) not displaying said phenotype, to form heteroduplex DNA containing a mismatch region at the site(s) at which the sequence of said first DNA diverges from the sequence of said second DNA;

b) cleaving one strand of said heteroduplex DNA to form a single-stranded gap across said mismatch region;

c) treating said cleaved heteroduplex with a DNA polymerase in the presence of dideoxynucleotides to determine the nucleotide sequence across said gap;

d) preparing a synthetic oligonucleotide comprising all or part of said nucleotide sequence; and

e) identifying a DNA clone that hybridizes to said oligonucleotide.

22. The method of claim 21 or 35, wherein the mismatch region is caused by one or more modifications in the gene of interest selected from the group consisting of additions, deletions, and substitutions of one or more nucleotides and combinations thereof.

23. The method of claim 21 or 35, wherein said nucleotide sequence is determined by enzymatic DNA sequencing.

24. A method for identifying one or more genetic alterations in a target sequence present in a first DNA sample, which comprises:

a) hybridizing said first DNA sample with a second DNA sample, wherein said second sample does not contain the alteration(s), to form heteroduplex DNA having free ends and containing a mismatch region at the site of an alteration(s);

b) cleaving said heteroduplex DNA at or in the vicinity of the alteration, forming new ends;

c) ligating a single-stranded oligonucleotide of predetermined sequence said new ends;

d) determining the nucleotide sequence of said DNA sample adjacent to said ligated oligonucleotide; and

e) comparing said nucleotide sequence with a predetermined cognate wild-type sequence to identify said genetic alteration(s).

25. The method of claim 24 or 35 further comprising blocking said free ends on said heteroduplex DNA prior to the cleaving step.

26. The method of claim 25, wherein the blocking step comprises a method selected from the group consisting of removal of 5' phosphate groups, homopolymeric tailing of 3' ends with dideoxynucleotides, and ligation of modified double-stranded oligonucleotides.

27. The method of claim 24 or 35, wherein said cleaving step comprises the steps of:

a) exposing said heteroduplex DNA to one or more non-protein chemical reagents under conditions appropriate for mismatch recognition and modification; and

b) cleaving one strand of said heteroduplex DNA in the vicinity of the modification.

28. The method of claim 27, wherein the chemical reagent is selected from the group consisting of hydroxylamine and osmium tetroxide.

29. The method of claim 24 or 35, wherein the single-stranded oligonucleotide is from about 15 to about 35 nucleotides in length.

30. The method of claim 24 or 35, wherein the ligating step is achieved using RNA ligase.

31. The method of claim 24 or 35, wherein the determining step is achieved using hybridization to oligonucleotide arrays.

32. A method for identifying one or more genetic alterations in a target sequence present in a first genomic DNA sample, the method comprising:

a) immobilizing said first DNA sample on a solid support;

b) hybridizing said immobilized sample with a second DNA sample, wherein said second sample does not contain the alteration, to form heteroduplex DNA having free ends and containing a mismatch region at the site of the alteration(s);

c) chemically blocking said free ends with a terminal transferase in the presence of a dideoxynucleotide;

d) cleaving one strand of said heteroduplex DNA adjacent to said mismatch region with bacteriophage T4 endonuclease 7 to form new ends;

e) ligating a single-stranded oligonucleotide having the sequence 5'-CAGTAGTACAACTGACCCTTTTGGGACCGC-3' to said new ends;

f) determining the nucleotide sequence adjacent to said ligated oligonucleotide using enzymatic DNA sequencing; and

g) comparing said nucleotide sequence with a predetermined cognate wild-type sequence to identify the mutation(s).

33. A method for identifying one or more mutation(s) in a DNA, the method comprising:

a) immobilizing said DNA sample on a solid support;

b) hybridizing said immobilized sample with a second DNA, wherein said second sample does not contain the mutation(s), to form heteroduplex DNA having free ends and containing a mismatch region at the site of a mutation(s);

c) chemically blocking said free ends;

d) cleaving one or both strands of said heteroduplex within or adjacent said mismatch region to form new ends;

e) ligating a single-stranded oligonucleotide of predetermined sequence to said new ends;

f) determining the nucleotide sequence adjacent to said ligated oligonucleotide; and

g) comparing said nucleotide sequence with one or more predetermined cognate wild-type sequences to identify said mutation(s).

34. A method for multiplex identification of one or more mutation(s) in a first DNA, the method comprising:

a) immobilizing one or more first DNA samples on a solid support;

b) hybridizing said immobilized sample(s) with a second DNA sample, wherein said second sample does not contain the mutation(s), to form heteroduplex DNA having free ends and containing a mismatch region at the site of a mutation(s);

c) chemically blocking said free ends;

d) cleaving one or both strands of said heteroduplex within or adjacent to said mismatch region, to form new ends;

e) ligating a single-stranded oligonucleotide of predetermined sequence to said new ends;

f) determining the nucleotide sequence adjacent to said ligated oligonucleotide; and

g) comparing said nucleotide sequence with one or more predetermined cognate wild-type sequences to identify said mutation(s).

35. A method for positional cloning of a gene of interest, the method comprising:

a) hybridizing a first DNA sample derived from an individual displaying a given phenotype with a second DNA sample, wherein said second sample is derived from one or more individual(s) not displaying said phenotype to forms heteroduplex DNA having free ends and containing a mismatch region at the site at which the sequence of said first DNA sample diverges from the sequence of said second DNA sample;

b) cleaving one or both strands of said heteroduplex DNA within or adjacent to the mismatch region to form new ends;

c) ligating a single-stranded oligonucleotide of predetermined sequence to said new ends;

d) determining the nucleotide sequence adjacent to said ligated oligonucleotide;

e) preparing a synthetic oligonucleotide comprising all or part of said nucleotide sequence; and

f) identifying a DNA clone that hybridizes to said oligonucleotide.

36. The method of claim 21 or 35, wherein the identifying step is achieved using a method selected from the group consisting of colony hybridization, identification of tissue specific expression, reverse transcription-amplification of mRNA, and screening of an affected population for genotype/phenoytpe association.